DERIVING FIXED BOND INFORMATION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of United States Provisional

Application Serial No. 60/120,614 entitled REPRESENTING A CHEMICAL

5 STRUCTURE filed on February 18, 1999, incorporated herein.

Background of the Invention

This application relates to deriving fixed bond information.

Some chemical structures may be depicted with circles or arcs instead of

with alternating single and double bonds (see the leftmost structure of Fig. 1).

10 Since a well known chemist named Auguste Kekulé suggested that double bonds

were not fixed or localized, and depicted a benzene ring with a delocalizing

circle instead of alternating single and double bonds, the process of perceiving

aromaticity (i.e., cyclic delocalization) from a fixed-bond representation may be

called "Kekulization." A Kekulé structure representation is one in which the

15 alternating single and double bonds of the classical depiction are replaced by all

single bonds, adorned by a circle or arc.

Summary of the Invention

A method and a system are provided for deriving fixed bond information.

In particular, a fixed bond representation of a chemical structure is derived from

20 a delocalized representation. A path is conceptually traced through the

represented structure and an examination is conducted, for each atom, of each

1

possible electronic and bonding state that is consistent with what has come before along the path. A result is found by examining states and orders in a semi-recursive procedure that is directed early towards likely answers. If there is more than one possible solution, the best solution is chosen by use of a rating function.

The method and the system are able to handle acyclic as well as cyclic systems, organic and organometallic structures, and can produce useful results in situations involving wide ranges of ring sizes and ring systems, element types, and hybridization. All or nearly all aromatic (i.e., cyclically delocalized) systems, including such systems with hetero substituents, and cyclic systems, and mixed cyclic-acyclic systems, can be addressed effectively by the method and the system. Charged systems, and systems with unpaired electrons can be handled as well. The method and the system are able to enumerate the mesomers of a structure.

Other features and advantages will become apparent from the following description, including the drawings, and from the claims.

Brief Description of the Drawings

Figs. 1-6, 8, 10, 12, and 15-21 are illustrations of output produced by software.

Figs. 7, 9, 11, and 13-14 are illustrations of computer data.

Figs. 22, 23A-23B, 24, 25A-25C, 26A-26B, 27-30, and 31A-31C are flow diagrams of computer-based procedures.

Detailed Description

A procedure is described in detail below in which a cyclic (i.e., aromatic) or acyclic delocalized system or structure diagram is dekekulized, i.e., in which a representation having alternating single and double bonds is derived from the delocalized system (see Fig. 1). In particular, a fixed-bond ("localized-bond") representation of an organic or organometallic structure is derived from a delocalized-bond representation of the structure.

The procedure allows fixed bond representations to be made more available to readers who prefer such representations, particularly when the pertinent bonding is part of a reaction that affects the pi system. (A pi orbital is the simplest atomic orbital that is amenable to delocalization. A pi system is a collection of pi orbitals that share electrons. As used herein, "pi" also refers to d, f and any other atomic orbitals that pool electrons with at least two other atoms.)

The procedure allows mesomers to be enumerated (see, e.g., Fig. 2). A mesomer is also known as a resonance structure. Mesomers are enumerated for at least two purposes: to understand delocalization and its relation to fixed bond representations and to rationalize or predict chemical reactivity.

Furthermore, computer programs that predict products may need to supply bond fixation information. A carbene example is shown in Fig. 3 which

3

demonstrates that a reaction of a pi system may entail bond fixation: all the fixed

bond mesomers of toluene are enumerated; each undergoes reaction; and

duplicate products, if any, are eliminated. With respect to hydrogenation, if an

aromatic or delocalized acyclic system is provided, and it is known that a

5   particular portion of hydrogen has been added, dekekulization is the first step in

locating various alternative locations for the added hydrogens and the residual

double bonds (see Fig. 4 which illustrates an application of dekekulization to the

assignment of hydrogenation isomers).

The procedure also allows determination of heteroatom hybridization and

10   the number of "implicit" hydrogens attached to heteroatoms. If hydrogen atoms

have been omitted from a drawing of a delocalized system (see, for example, the

lefthand structure of Fig. 1), the procedure reveals in many or most cases the

hydrogen count at each atom necessary to make chemical sense of the structure.

As suggested in Fig. 5, this information is not generally apparent from simple

15   graph theory. Fig. 5 illustrates the unpredictability of heteroelement

hybridization, and resulting implicit hydrogen count. It may be desirable to

generalize from pyrrole and pyridine that heteroatom hybridization follows

from ring size, but such a generalization can prove unreliable, as demonstrated

by the two counterexamples at right in Fig. 5.

20   Inferring fixed bond orders from a delocalized representation is also an

important process in computer-assisted organic synthesis.

4

One or more of the procedures related to dekekulization as described
herein may be used in or with one or more procedures described in the following
United States patent applications, which are incorporated herein: DERIVING
CHEMICAL STRUCTURAL INFORMATION, serial no. _____, filed February 11,

5   2000, and ENHANCING STRUCTURE DIAGRAM GENERATION, serial no.
_____, filed February 11, 2000. For example, dekekulization as described herein
may be used to help derive a chemical structure diagram from a chemical name.

Conceptually, in a specific embodiment of the procedure, a path is traced
through the structure and, for each atom, each possible electronic and bonding

10   state is examined that is consistent with previous results along the path. By
extensively or exhaustively examining possible states and orders, the procedure
is able to arrive at a fixed bond solution, if one exists.

An inspection of fixed-bond structure diagrams reveals, with respect to
carbon atoms, that a pair of adjacent delocalized bonds tends to be replaced by

15   one single and one double bond. If a carbon atom stands as a fusion atom, e.g.,
as one of the tertiary atoms in naphthalene, the third bond is single, and may be
viewed as an appropriation of the fourth bond that otherwise would be directed
outside of the delocalized system. Nitrogen, on the other hand, typically may
have either two single bonds directed into the system plus a single bond

20   externally directed, as in pyrrole, or a single and a double directed into the
system, with no external bonds, as in pyridine.

5

In a specific implementation, in the procedure, every atom is described by a three-digit code representing a bonding environment. In many cases, carbon's code is [121], which indicates that of the two originally delocalized bonds, one becomes single and the other becomes double, and that there is one residual

5   valence unit to be deployed either outside the system (e.g., a sigma bond) or with a third atom within the system (naphthalene). The code can be enhanced to indicate the number of pi electrons formally contributed to the delocalized system by an atom in the pertinent state. For the common carbon state, the number is one, indicated by the enhanced code [121/1].

10   Information for an atomic environment may refer to a charge or an unpaired electron or both, in addition to the bonding information. The collection of characteristics involved is referred to hereinafter as an environment or as an electronic state and valence distribution ("ESVD"). A carbon atom as commonly encountered in organic chemistry, with two attachments that are part of the

15   delocalized system, is found only in the environments shown in Fig. 6. In Fig. 6, which illustrates the electronic/bonding environments available to a carbon atom with two delocalized attachments, dashed bonds represent the fixed bonds that are replacing delocalized bonds. (Although neutral and fully paired, the rightmost ESVD tends to be a poor choice due to poor overlap of its pi orbital

20   with the two delocalized attachments of interest.) Common codes for nitrogen are [111/2] (pyrrole) and [120/1] (pyridine). Prevalent atomic environments and

6

corresponding shorthand codes of the form [WXY/Z] are listed in Table 1 (Fig. 7). The atomic ESVD environments listed in Table 1 are illustrated in Fig. 8, in which dashed bonds represent the fixed bonds that are replacing delocalized bonds.

5   Table 1 is not a complete list, but furnishes the essential elements and values for at least some cases. With respect to the external bonds column, if an atom has a third delocalized bond, an "external bond" is appropriated to represent it. With respect to the text "like N" in the first entry for element P, a "treat-like" element (such as P) inherits all of the entries for its parent (such as N

10   in the case of P) and may have additional unique entries.

All possible ESVDs may be attempted for all atoms, and all bond orders (including single and double) may be attempted for all bonds. A solution is determined to have been found when a combination of codes and orders is self-consistent, such that, for every atom, the orders of bonds to the atom match the

15   requirements of the atom's ESVD, and the net charge and number of paired electrons of the system are also as required by the structure. Possible further requirements include a satisfactory 4n+2 electron count or an absence of radicals or zwitterions.

In a specific implementation, the procedure includes several practical

20   features that help to produce timely results in practice. A first of the practical features reflects a recognition that it is not always necessary to try all ESVDs and

bond orders: with respect to choosing an order for a bond to an atom having an assigned ESVD, the only orders considered are orders that are consistent with the ESVD, taking account of bonds already assigned. For example, if a carbon ([121/1]) has been assigned a double bond, the carbon's next bond may not be

5   double, and is only considered to be single. Further, the only ESVDs that are chosen are ESVDs that are consistent with adjacent fixed bonds.

A second of the practical features is consistent with a recognition that in practice many or most solutions do not involve unpaired electrons or charge: ESVDs featuring unpaired electrons or charge are not considered initially. If a

10   solution is found without referring to such ESVDs, the procedure is finished, and time has been saved.

According to a third of the practical features, the procedure is only partially recursive. With respect to bond orders, when alternative bond orders are attempted, recursion is sensible: if a single bond appears to be the most

15   advantageous next step, and its recursive development returns in failure, the double bond remains the most advantageous next option. In other words, pursuit of a bond order is exhaustive. By contrast, with respect to ESVDs, some ESVDs tend to be more promising (i.e., better) than others, in at least some cases. Accordingly, after the possible ESVDs for an atom are tabulated, the best of the

20   possible ESVDs is actually pursued, recursively, and the rest of the possible ESVDs are placed in a priority queue, keyed to a rating of the inchoate structure's

8

likelihood of success. In this way, less promising ESVDs are not examined until more promising combinations are exhaustively considered.

Another of the practical features in a computer-based implementation is an implementational measure that speeds the assessment of compatibility

5   between bonds and ESVDs: a screening bitmask. (A bitset is a piece of data, e.g., an integer, in which each bit represents a Boolean value. As used herein, a bitmask is synonymous with a bitset.) For a given ESVD, a screening bitmask encodes the ESVD's bonding and electron requirements. Similarly, for an atom, another screening bitmask specifies which types of bond have already been fixed.

10   When the latter screening bitmask is logically (i.e., bitwise) subtracted from the former screening bitmask, a list of bond types still required at the atom is obtained. By logically subtracting the former screening bitmask from the latter screening bitmask, it can be determined whether it is possible to apply the ESVD to the atom. Such operations tend to save time, since multiple logical

15   comparisons are compressed into one comparison of bitmasks.

In a specific implementation, the screening bitmask for the ESVD [121], for example, can be figured from Table 3 (Fig. 11). Only the bonding portion of the bitmask is considered when choosing the bond orders compatible with an ESVD. The internal single bond, the internal double bond, and the external single bond

20   of the ESVD [121] correspond to bits #0, #2, and #4, respectively. Therefore, the bitmask has the value {0,2,4}. For an atom that has had one single bond fixed, the

bitmask consists of bit #0, with the value {0}. The difference between the former and latter bitmasks, {0,2,4} - {0} = {2,4}, reveals the fixed bond types that must eventually be assigned to the atom, namely, an internal double bond (bit #2 and an external single bond (bit #4).

5   An example of the opposite process is determining whether an ESVD, such as [111/2-], is compatible with an atom that, for example, has two fixed single bonds and which, due to a global restraint, may not be charged. The ESVD's bitmask is {0,1,4,10,12,14}, indicating that the ESVD has an internal single bond, has two internal single bonds, has an external bond, is negative, is not positive,

10   and does not have an unpaired electron, respectively. The atom's bitmask is {0,1,8,11,12}, indicating the atom has an internal single bond, has two internal single bonds, is neutral, is not an anion, and is not a cation, respectively. The atom's bitmask {0,1,8,11,12} is not a subset of the ESVD's bitmask {0,1,4,10,12,14}, which indicates that the ESVD is not compatible with the atom. Although the

15   bonding portions are compatible, the electronic portions are not. (In general, the more fully characterized or developed an atom is, the more bits are set in its bitmask, and the less likely that the atom will be compatible with a given ESVD.)

Different applications of the dekekulization procedure may have slightly different requirements. Thus, in at least some cases, it may be advantageous for

20   the procedure to refer to a bitset of control flags, listed in Table 2 (Fig. 9).

An example 1000 of the procedure is illustrated generally in Figs. 10, 22 and is now discussed generally; a more detailed example is included below. Fig. 10 illustrates an example of environment development and backtracking for furan. In this particular example, for simplicity, solutions that incorporate

5   charges or unpaired electrons are not considered. Initially, none of the atoms or bonds has been assigned. An initial atom is selected (step 1010). For example, the top left atom is arbitrarily selected and is assigned the only neutral and non-radical environment for carbons that is also free of external pi bonding: [121/1] (see Fig. 6). Thus the top left atom will have a single and a double bond disposed

10   within the originally delocalized system, and a single bond disposed without. The top left atom contributes one pi electron to the delocalized system.

A bond order is assigned to an adjacent bond (step 1015). Either a single bond order or a double bond order may be selected. In this case, a double bond order is arbitrarily chosen at this point. As explained below, after the

15   ramifications of the double bond order choice have been extensively or exhaustively explored, the procedure returns to this point and proceeds with the single bond order choice instead.

A next adjacent atom is assigned an ESVD that is consistent with the previous bond, here, the double bond (step 1020). Since the example includes no

20   ions or radicals, the ESVD that is assigned is [121/1].

11

The next adjacent bond's order is selected (step 1030). Since the preceding atom has a double bond, the only choice for the order that is consistent with the atom's environment is a single bond.

The same approach is repeated for additional atoms (step 1040).

5   If an unacceptable state is encountered, the path values are rejected (step 1050). In this case, a state is encountered in which the oxygen atom has one double bond and one uncharacterized bond. Since no (ESVD) type of uncharged oxygen can accommodate this state, the state is unacceptable and the path values are rejected.

10   The procedure backtracks to the last point where a selection was made, and proceeds forward from there with a different selection (step 1060). In this example, the procedure proceeds forward with the single bond order choice as noted above.

If a solution is found that cannot be improved upon, i.e., if all atoms and

15   bonds have been assigned mutually compatible environments and orders, respectively, and the collection is holistically optimal, the procedure terminates (step 1070). On the other hand, if a solution is found that is not optimal and there are other choices, such as were encountered with respect to the alternative bond order choices above, the other choices are explored (step 1080). The

20   principal factors in the holistic value of the solution are the count of the

12

delocalized electrons in each ring, and whether an excessive number of charges or radicals or both are present.

A more detailed example of the procedure is now described, with the use of terms as now explained. A "delocalized atom" is an atom having a vacant pi orbital that is shared, or one or more pi electrons that are shared, with one or more adjacent atoms. (The term pi is intended to encompass d and f orbitals and any other orbitals that participate in delocalized bonding.) A bond that holds the one or more shared pi electrons is called a "delocalized bond", and in the delocalized style of depiction is drawn as a single bond with a nearby arc or circle.

The largest set of contiguous delocalized bonds containing a given delocalized bond is called a "delocalized system". The atoms in the delocalized system are those adjacent to the delocalized bonds. As a result, an atom or bond may belong to at most one delocalized system. Contributions emanating from inside and outside the delocalized system are "internal" and "external", respectively. For example, sigma bonds and *exo* double bonds are external.

The "internal coordination number" ("ICN") of a delocalized atom is the count of its adjacent atoms that belong to the same delocalized system. Although only ICNs from 1 to 3 are described herein, the concepts involved are extensible to larger ICNs.

13

A delocalized atom's "internal order" is its bond order sum, after bond fixation, contributed by formerly delocalized bonds, and cannot be determined before bond fixation. A delocalized atom's "external" order is the bond order sum emanating from outside the delocalized system, and is known from the outset unless implicit hydrogen atoms are in question.

Each element is associated with a respective list of electronic state and valence distribution ("ESVD") values that are available (see Table 1 in Fig. 7). An ESVD specifies the number of bonds to an atom, their orders, and the charge and radical state of the atom. Separate accounting is made for internal and external bonds. Additionally, a distinction may be made between an external multiple bond and a set of several external bonds of lower order, where the sum of the bond orders in the set equals the order of the external multiple bond. A solution is found when mutually consistent ESVDs (i.e., with correct bonding) have been found for all atoms, and the net charge and radical count are correct for the system.

An ESVD screening bitmask ("ESB") is used to determine whether a given ESVD is compatible with a given atom. The ESB is a constant for a given ESVD, and is computed from the corresponding ESVD at run time (see procedure 8000 in Fig. 29, discussed below). The ESB is a bitset (i.e., a bitmask), which is a binary number, each bit of which represents a Boolean value indicating whether a

14

condition is present or absent. Table 3 (Fig. 11) illustrates the meanings of the bits in the ESB.

An atom screening bitmask ("ASB") has the same bit values as the ESB (Table 3, Fig. 11), but is computed slightly differently (see procedure 9000 in Fig.

5   30, discussed below). An ESVD can be assigned to a given atom if the atom's ASB is a subset of the ESVD's ESB.

A "path" taken through a system (cyclic or not) is a type of directed graph, and is represented by a vector of atom and bond numbers, in the order in which the atoms and bonds will be considered. Fig. 12 illustrates an example of a

10   path, marked with atom and bond (underlined) numbers, in which ring closure atoms appear in bold; the path may be discontinuous. It is not necessary for consecutive path bonds to be adjacent, as long as every later-occurring bond is adjacent to any of the earlier-occurring bonds.

A "strategy" for the system is a script that drives path development, and

15   is formulated at the same time the path is created. Various actions possible in a step of the script are shown in Table 4 (Fig. 14). The strategy for the example of Fig. 12 is shown in Fig. 13.

A "state" is a data structure that includes the information about the choices made in the path up to the current point, and refers to a "residual

20   charge", a "residual radical", and a "strategy step index", as well as the ESVD and bond order assignments. A residual charge is the amount of charge remaining to

be dispelled by a suitable choice of charged atoms. For example, a residual charge of +2 requires that two cations, or three cations and an anion, or other ion set that adds up to a +2 charge, be chosen in the course of the rest of the path. A residual radical can be zero ("no-radical") or one ("radical"). If the residual

5   radical is one, an odd number of radicals must be assigned in the course of the rest of the path. If the residual radical is zero, an even number must be so assigned. A strategy step index notes the most recent step number executed in the strategy list.

In at least some cases, one or more of the features presented may be

10   advantageous but not essential. For example, the pre-identification of the path and the use of a strategy script based on the path, are logical simplifications; instead of iterating over the script's steps, the part of the procedure that assigns a bond order or an ESVD may directly invoke the next logically succeeding step. In at least some computer-based implementations, maintenance of the strategy

15   involves processing overhead that is not insignificant, but that also contributes logical order and stability. Similarly, the use of an ESB or an ASB may be an implementational convenience that is not essential. Note that all quantitative values cited herein are provided for illustrative purposes only, and are not necessarily optimal for any or all cases.

20   In an example implementation, the dekekulization procedure includes a procedure 2000 (Figs. 23A-23B), now described, that is provided with

16

information describing a chemical structure, an indication of which of the structure's bonds are of interest, a bitmask of control flags, and an indication of a timeout period. In at least some cases, the bonds of interest include the complete set spanning the structure. The procedure ignores non-delocalized bonds. The

5   bitmask of control flags is as described above (Table 2 in Fig. 9). The timeout period is the interval after which the procedure should terminate and present its best solution, even if that solution is imperfect. In a specific implementation, the timeout period is more than sufficient in a typical case, for which the production of the fully satisfactory solution is nearly instantaneous. Procedure 2000 returns

10   an indication of success or failure, depending on whether bonds have been successfully fixed.

In procedure 2000, discrete (i.e., not adjacent) delocalized systems present in the structure are identified (step 2010). (As used herein, "structure" refers to a collection of one or more molecules, and an ion qualifies as a molecule.) Each

15   next system is identified by selecting any delocalized bond that has not already been treated (step 2020), adding all adjacent delocalized bonds (step 2030), and continuing to add delocalized bonds adjacent to those just found, until no new bonds are added (step 2040).

Steps 2060 - 2120, now described, are executed for each of the identified

20   delocalized systems (step 2050). The instant delocalized system is denoted "*DS*".

DS is analyzed for characteristics (step 2060). In a specific implementation, the

17

analysis proceeds as follows. The total charge and radical count required of the pi system are noted (step 2070). (Charges localized in sigma orbitals, such as in phenyl lithium, are not included, and are ignored throughout procedure 2000.) The internal coordination number ("ICN") of each atom is calculated (step 2080).

5 Bonds in *DS* are identified that must be fixed, as a consequence of having an adjacent fixed multiple bond (step 2090). (This is limited to carbon atoms if there is ambiguity in the valence of polyvalent elements.)

A path through *DS* is developed (step 2100). In at least some cases, it does not matter substantially whether the path is breadth-first, depth-first, or is of

10 some other type. However, it has been found in some cases that highly efficient paths tend to be those that complete rings promptly. The path need not be continuous, although every path atom (bond) other than the first must be adjacent to one previously visited. If any fixed bonds were identified in step 2090, the process begins with an adjacent atom, which tends to limit the choices

15 for the succeeding bond order. (In general, in at least some cases, it is advantageous to defer choices in state and bonding to as late in the path as possible.)

The strategy list is constructed (step 2110). The strategy list includes directives to assign ESVDs and bond orders for the atoms and bonds encountered

20 in the path, and additionally directives to check each atom's final environment the last time the atom is visited.

Ring systems are identified (step 2120). A ring system is a group of one or more rings, each of which shares one or more bonds with another in the group, that cannot be divided into two or more smaller ring systems without breaking a ring bond. As used herein, "ring" refers only to a cycle that contains one or more

5   delocalized bonds, or that is adjacent to such a cycle. Other cycles are ignored.

The best solution ("BS") is set to "undefined" (step 2130).

For each of three modes, now described, steps 2150 - 2190 described below are executed for as long as they do not terminate, either as a result of an identification of a solution deemed perfect, or by exhausting possibilities, and the

10   time elapsed does not exceed an allocated amount of time (step 2140). Any mode that is incompatible with the control flags is skipped (step 2150). According to mode 1, charges or radicals beyond those provided are not created, and any charges (and optionally radicals) are confined to heteroatoms. In the case of mode 2, charges and radicals are handled in the same way as in mode 1, except

15   that there is no restriction on their placement. In mode 3, radicals and charges are freely created, without restriction on their placement.

The state is initialized (step 2160) as follows. The state's charge count is set to that of the system. If the system has an odd number of unpaired electrons, the residual radical flag is set; otherwise it is cleared. The state's strategy step index

20   is set to zero, indicating a point immediately preceding the first step in the strategy.

19

The state is queued (the queue is effectively empty at this point) (step 2170).

While there are queued states, the most promising ("S") is dequeued and is passed to a dispatcher procedure ("Dispatcher", described below), (step 2180).

5   If the Dispatcher returns *"Perfect Solution"*, the procedure skips to step 2210.

If *BS* is undefined after the mode processing is complete, an indication of failure is returned, and the entire procedure terminates (step 2200).

BS is applied to the given structure (step 2210), so that bond orders are fixed and charges and radical values are assigned to atoms as required by *BS*.

10   (Where identification of implicit hydrogens is of interest, these are determined from the ESVDs of the heteroatoms. That is, if the valence of a heteroatom's ESVD is larger than the its apparent valence after bond fixation, the difference is made up with hydrogen atoms. The entire procedure terminates, returning a value of success.)

15   The Dispatcher executes as follows when provided with a state ("S"). In general, the Dispatcher recursively executes the next strategy step and returns either *"Perfect Solution"* or *"Keep Trying"*. In a specific implementation, the following steps are executed (procedure 3000, Fig. 24). If the time elapsed has exceeded the amount of time allocated, *"Keep Going"* is returned (step 3010).

20   Otherwise, the strategy step index for S is incremented (step 3020), and the Dispatcher continues as follows. The procedure ("action") associated with the

instant strategy step is executed (see Table 4 in Fig. 14), as indicated below (step

3030), and if the respective procedure returns *"Perfect Solution"*, the Dispatcher

returns *"Perfect Solution"* (step 3040). Procedures 4000-7000 are described below.

Assign ESVD:	Execute procedure 4000.
Assign Bond:	Execute procedure 5000.
Verify Completed:	Execute procedure 6000.
Complete:	Execute procedure 7000.

If the respective procedure does not return *"Perfect Solution"*, the strategy step

index for S is decremented (step 3050), and *"Keep Trying"* is returned (step 3060).

10   In procedure 4000 (Figs. 25A-25C), an ESVD is assigned. Generally,

procedure 4000 is provided with the state (*"S"*), as so far developed, and an atom

(*"a"*) to be assigned an ESVD, and returns either *"Perfect Solution"* or *"Keep

Trying"*. Zero or more of *a*'s bonds have been fixed at this point. In a specific

implementation, the following steps are executed.

15   A bitmask BM_a that describes *a* is formed using procedure 9000 (Fig. 30)

(step 4010) For each of *a*'s fixed bonds, the bit is set in BM_a corresponding to its

bond order (see Table 3 in Fig. 11). If *a* has an external bond, the corresponding

bit is set in BM_a. If the operative mode does not permit spontaneous generation

of radicals, the no-radical bit is set. If the operative mode does not permit

20   creation of charges, the no-plus and no-minus bits are set.

The ESVDs available to a include all of the element's ESVDs that contain, bitwise, all of the bits present in Bm_a.

A set ("C") is determined of ESVDs of a's element type that are compatible with a (step 4020). An ESVD is determined to be compatible if the ESVD has an ESB that is a superset of BM_a. If no ESVDs qualify for inclusion in C, "Failure" is returned (step 4030).

A penalty ("P") is assigned to each member (e) of C (step 4040). The penalty represents an assessment of the strategic desirability of using the ESVD. The penalty value is zero or positive, with larger values corresponding to lower desirability. In a specific implementation, each penalty is calculated as follows (steps 4050 - 4110). Initially, the penalty (P) is set to zero (step 4060). If the ESVD represents a radical, and the state's residual radical flag is clear, P is incremented by three (step 4070). If the ESVD is charged, and the residual charge of the state is non-zero and of opposite sign, 50 is added to P, thus practically eliminating the ESVD from consideration (step 4080). If the ESVD is charged but the state's residual charge is zero, P is incremented by 2 (step 4090). If e (i.e., the ESVD) is charged and a is carbon, P is incremented (step 4100). If e does not have an internally directed multiple bond, P is incremented (step 4110).

At this point, each penalty has been calculated. The members of C are sorted in order of increasing penalty (step 4120).

Inferior ESVDs are queued for later consideration (step 4130). In a specific implementation, the queuing proceeds as follows (steps 4140 - 4180). For each member (C_i) of C except the first (i.e., the best), the following steps 4150 - 4180 are executed (step 4140). The current state is copied to a new state, q (step 4150).

5   The ESVD of atom a in q is set to C_i (step 4160). The strategy step counter for q is incremented, so that when the state is dequeued, state development will resume after a's ESVD assignment (step 4170). Further, q is inserted into the priority queue, and is ranked therein based on the penalty respectively assigned as described above (step 4180). (In a case in which the queue has a finite capacity,

10   the worst entry is deleted to make room for the new one.)

At this point, the inferior ESVDs have been queued. The best ESVD is applied to S (step 4190). In a specific implementation, the application is executed as follows (steps 4200 - 4230). The state variables having to do with atomic properties are updated (step 4210): the ESVD's charge is subtracted from the

15   state's residual charge (step 4220), and if the ESVD has an unpaired electron, the state's residual radical flag is toggled (step 4230).

The Dispatcher (i.e., procedure 3000) is recursively invoked (step 4240), and if the Dispatcher returns "*Perfect Solution*", the instant procedure returns the same (step 4250).

20   Otherwise, S is restored to its value before the best ESVD was applied (step 4260), and "*Keep Trying*" is returned (step 4270).

23

In procedure 5000 (Figs. 26A-26B), a bond order is assigned. Procedure 5000 is provided with the current state ("S"), the bond to be fixed ("b"), and the atom ("a") preceding this bond in the path, and returns *"Perfect Solution"* or *"Keep Trying"*. In a specific implementation, the bond order is assigned as follows

5 (steps 5020 -5150). The ASB of the preceding atom is calculated (step 5020) and is called BM_a. BM_{esvd} is the ESB of a's ESVD.

The difference $BM_{diff} = BM_{ESVD} - BM_a$ is calculated (step 5030). Only the bond-order bits (0-5 of Table 3 of Fig. 11) are of interest. The presence of a given bond-order bit in BM_{diff} indicates the particular option is available.

10 The possible bond orders are determined (step 5040). In a specific implementation, the determination proceeds as follows (steps 5050 - 5150). *AddSingle* is set to true if and only if BM_{diff} has either the *internal single* bit or the *two internal single* bit set (step 5050). *AddDouble* is set to true if and only if BM_{diff} has the *internal double* bit or the *two internal double* bit set (step 5060). *CooptExtern*

15 is set to true if a's ICN is three or greater, and a's valence allows a one or more units of bond order beyond the units already used up by a's fixed internal bonds (step 5070).

Fig. 15 shows an example of when an external bond is co-opted to serve as an additional internal single bond. In part a of Fig. 15, an atom has one fixed

20 single bond, and its ESVD (for example, [121]) has just been assigned. Ordinarily the order of the next bond (marked by the arrow) would have to be double, but

24

because the atom has three internal bonds, the order might also be single. In part b of Fig. 15, all of the nitrogen's bonds have been fixed but one. Normally no bond order could be chosen, because two internal bonds of the nitrogen's ESVD, [111], have been consumed. However, because the nitrogen's ICN is 3, the external bond may be reassigned as an internal (single) bond.

If either *AddSingle* or *CooptExtern* is set to true, *b*'s order is set to single and any relevant state variables are updated (step 5100), and the Dispatcher is recursively invoked (step 5110). If the Dispatcher returns *"Perfect Solution"*, the instant procedure returns the same (step 5120).

If *AddDouble* is set to true, *b*'s order is set to double, updating any relevant state variables (step 5130), and the Dispatcher is recursively invoked (step 5140). If the Dispatcher returns *"Perfect Solution"*, the instant procedure returns the same (step 5150).

Otherwise, the instant procedure returns *"Keep Trying"* (step 5150).

In procedure 6000 (Fig. 27), an atom is verified as complete. In particular, the procedure is provided with the state (*S*) and an atom (*a*), all the bonds of which have just been fixed. The procedure determines whether *a*'s assigned ESVD is compatible with its bond orders, and returns *"Perfect Solution"* or *"Keep Trying"*. In a specific implementation, the procedure executes as follows (steps 6020 - 6080). *a*'s ASB is calculated using procedure 9000 (Fig. 30) (step 6020) and is called BM_a. BM_{esvd} is set to the ESB of *a*'s assigned ESVD. Only the bonding bits (0-

25

5 of Table 3 of Fig. 11) of BM_a and BM_{esvd} are considered. BM_a is corrected for

tertiary atoms (step 6030), so that if BM_a equals {0,1,2} (see Table 3 of Fig. 11), BM_a

is set instead to {0,2} (step 6040).

If the ESVD calls for U units of external bonding, but the atom does not

5   express U units of external bonding, and the control flags permit implicit

hydrogens, bit 4 is included in BM_a (step 6050).

If $BM_a = BM_{esvd}$, the assigned ESVD is satisfactory. In this case, the

Dispatcher is recursively invoked (step 6060), and the instant procedure returns

whatever value the Dispatcher returns (step 6070).

10   Otherwise, the assigned ESVD is determined not to be viable, and "*Keep*

Trying" is returned (step 6080).

Procedure 7000 (Fig. 28), when provided with a state (S), which represents

a complete traversal of the path, determines whether S offers a perfect solution,

and returns "*Perfect Solution*"or "*Keep Trying*". If an imperfect solution is viable, it

15   is noted. In a specific embodiment, the determination is executed as follows

(steps 7020 - 7070). If S has a residual charge or radical, "*Keep Trying*" is returned

(step 7020).

At this stage, S constitutes a viable solution.

If mesomers are being enumerated, the dekekulized structure may be

20   offered, e.g., via a callback function, back to the client program or other source

that invoked dekekulization (step 7030).

S's rating (*R*) is calculated by use of procedure 10000 (Figs. 31A-31C) (step 7040). If *R* is zero, indicating a perfect solution, and mesomers are not being enumerated, "*Perfect Solution*" is returned (step 7050). If the best solution (*BS*, initialized in step 2130) is undefined, or if *R* is better than the rating of *BS*, *BS* is set to *S* (step 7060), and "*Keep Trying*" is returned (step 7070).

In procedure 8000 (Fig. 29), an ESB is calculated for an ESVD. Procedure 8000 is provided with an ESVD and returns the ESB (see Table 3 in Fig. 11 for bit descriptions). In a specific implementation, the calculation proceeds as follows (steps 8010 - 8060). ESB is initialized to zero (step 8010). The following bits of ESB are set according to the ESVD's charge (step 8020):

Neutral: {8,11,12}

Plus: {9,11}

Minus: {10,12}

The following bits are set according to the ESVD's radical character (step 8030):

Radical: {13}

Non-radical: {14}

The following bits are set according to the ESVD's pair of internal bond orders (step 8040):

Single, single: {0,1}

Single, double: {0,2}

27

Double, double: {2,3}

The following bits are set according to whether the ESVD has one or more

units of external bonding (step 8050):

Has an external bond: {4}

5   Does not have an external bond: {5}

The ESB is returned (step 8060).

In procedure 9000 (Fig. 30), an ASB is calculated. Procedure 9000 is

provided with an identification of an atom (a), some or all of the bonds of which

may have been fixed, and the current state (S), and returns a's atom screening

10   bitmask (ASB). (See Table 3 in Fig. 11 for bit descriptions.) In a specific

implementation, procedure 9000 executes as follows (steps 9010 - 9090). ASB is

initialized to zero (step 9010). The following bits of ASB are set according to the

number of internal bonds adjacent to a that are fixed, with bond order single

(step 9020):

15   One: {0}

Two or more: {0,1}

The following bits are set according to the number of internal bonds

adjacent to a that are fixed, with bond order double (step 9030):

One: {2}

20   Two or more: {2,3}

28

If a has more than two adjacent bonds, bit {4} is set (step 9040). If a has exactly two adjacent bonds, and if the control flags dictate that no implicit hydrogens exist, and a's element type is not carbon (since carbons are always permitted implicit hydrogens in this implementation), bit {5} is set (step 9050). If the residual radical value of S is "no-radical", and the control flags do not permit unnecessary creation of radicals, bit {14} is set (step 9060).

If the residual charge of S is zero and the control flags do not permit unnecessary creation of charged atoms; or if a's element type is carbon and the control flags require charges to be situated on heteroatoms, bits {11,12} are set (step 9070); otherwise, if the residual charge of S is positive, bit {11} is set.

If the residual charge of S is negative, bit {12} is set (step 9080).

The ASB is returned (step 9090).

In procedure 10000 (Figs. 31A-31C), a state is rated ("scored"). Procedure 10000 is provided with a state (S) that incorporates the entire path through the current delocalized system (DS), and returns an integer (R) that rates the quality of the solution. R is zero or negative, with zero reflecting a perfect solution, and less negative values being better than more negative ones. In a specific implementation, procedure 10000 executes as follows (steps 10010 - 10250). R is initialized to zero (step 10010).

For each ring system ("RB") in DS, the following steps 10030 - 10250 are executed (step 10020). Periphery (P), which is to become the set of bonds in the

periphery of the ring system, is set to "empty" (step 10030). For each ring r

belonging to the RB, the following steps 10050 - 10110 are executed (step 10040).

P is set to the exclusive OR of itself with the bonds in r (step 10050). (Taking the

exclusive OR of two sets produces a third set having the value zero where the

5   corresponding bits in the two given sets match, i.e., 0=0 or 1=1, and one

otherwise.) The "one-ring" penalty is assessed (step 10060) as follows (steps

10070-10080). Where e represents the number of pi electrons in r, e is the sum of

the "electrons contributed" value of the ESVD (see Table 1 in Fig. 7) for each atom

in r. If e is a multiple of four ("4n"), i.e., conventionally anti-aromatic, and the

10   control flags prescribe the penalization of anti-aromatic solutions, two is

subtracted from R (step 10070). If e is odd, one is substracted from R (step 10080).

For each ring r2 belonging to RB wherein r2 > r, the following steps 10100

- 10110 are executed (step 10090). If r and $r2$ have one or more rings in common,

RC is set to be the compound ring of r and $r2$, i.e., the ring obtained by taking the

15   exclusive OR of the bonds in r and $r2$ (step 10100) and R is adjusted by the one-

ring penalty amount assessed in steps 10060-10080, where $r2$ is substituted for r,

(step 10110).

If there are three or more rings in RB, $r3$ is taken to be the compound ring

represented by P (step 10120), and R is adjusted by the one-ring penalty amount

20   assessed in steps 10060-10080, where $r3$ is substituted for r (step 10130).

30

A significant penalty is applied for gratuitous charges and radicals (step 10140), as follows (steps 10160 - 10180). NSC is taken to be the number of superfluous charges, calculated from:

NSC = (number of cations in S + number of anions in S) - $|C|$

where $|C|$ is the absolute value of DS's net charge (step 10160). NSR is taken to be the number of superfluous radicals, calculated from:

NSR = number of radical atoms in S - Q

where Q is zero if DS has an even number of unpaired electrons, and one otherwise (step 10170). The score is reduced by ten times ($NSC + NSR$) (step 10180).

A small penalty is applied for locating a charge on a carbon rather than an available heteroatom (step 10190), as follows (steps 10200 - 10220). NCC is taken to be the number of charged carbon atoms in S (step 10200). NNH is taken to be the number of neutral heteroatoms in S (step 10210). If NCC exceeds NNH, R is reduced by ($NCC - NNH$) (step 10220).

Pairs of adjacent atoms having identical, either empty or full, orbitals are penalized (step 10230). For example, in at least some cases, it is not desirable for two [111/0] boron atoms, or two [111/2] (pyrrole-type) nitrogen atoms, to be adjacent. For each pair of adjacent atoms in DS, five is substracted from R if the atoms in the pair have the same pi electron contribution (see Table 1 in Fig. 7) which is not one.

31

R is returned (step 10250).

Figs. 16-20 illustrate examples related to the procedures described above. Figs. 19 and 20 illustrate radical structures and acyclic examples, respectively.

All or a portion of the procedures described above may be implemented

5   in hardware or software, or a combination of both. In at least some cases, it is advantageous if the technique is implemented in computer programs executing on one or more programmable computers, such as a personal computer running or able to run an operating system such as UNIX, Linux, Microsoft Windows 95, 98, 2000, or NT, or MacOS, that each include a processor, a storage medium

10   readable by the processor (including volatile and non-volatile memory and/or storage elements), at least one input device such as a keyboard, and at least one output device. Program code is applied to data entered using the input device to perform the technique described above and to generate output information. The output information is applied to one or more output devices such as a display

15   screen of the computer.

In at least some cases, it is advantageous if each program is implemented in a high level procedural or object-oriented programming language such as Perl, C, C++, or Java to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. In

20   any case, the language may be a compiled or interpreted language.

In at least some cases, it is advantageous if each such computer program is stored on a storage medium or device, such as ROM or optical or magnetic disc, that is readable by a general or special purpose programmable computer for configuring and operating the computer when the storage medium or device is

5   read by the computer to perform the procedures described in this document. The system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner.

10   Other embodiments are within the scope of the following claims. For example, one or more of the procedures described above may be applied to structures other than common organic structures. For example, benzynes and other systems, cyclic and acyclic, contain one or more bonds involving pi orbitals that are not delocalized. These cases may be treated by formally assigning the pi

15   bond to the external bonding network, as if each of the pair of atoms had an external attachment instead of an internal bond. In another example, charges localized in a sigma orbital (as in PhLi) may be ignored, as may be localized radicals. Another example involves multi-center bonds ("μ bonds"). Many organometallic compounds, such as ferrocene and allyl copper (see Fig. 21 which

20   illustrates multi-center bonds), are depicted in the delocalized representation with a bond between an atom, usually a metal, and a pi network, rather than

33

another atom. Effectively, no aesthetic representation exists for these structures that eliminates the multi-center bond. More particularly, in at least some cases, in order to fix the bonding in the delocalized system (cyclopentadienyl, ally), either charge or unpaired electrons or "exhaustive sigma" bonds must be

5   introduced.